The component counts of random injections

Dudley Stark
School of Mathematical Sciences
Queen Mary, University of London
London E1 4NS, UK

Submitted: Apr 5, 2020; Accepted: Dec 6, 2020; Published: TBD

Abstract

A model of random injections is defined which has domain $A \cup B$ and codomain $A \cup C$, where A, B and C are mutually disjoint finite sets such that $|B| \leqslant |C|$. The model encompasses both random permutations, which is the case $B = C = \emptyset$, and random maximum matchings of a complete bipartite graph, which is the case $A = \emptyset$. The possible components of random injections are cycles and paths. Results on the counts of cycles and paths of different sizes are obtained for this model.

Mathematics Subject Classifications: 05C20, 05C30, 05C80

1 Introduction

Suppose a permutation is chosen uniformly at random from the symmetric group S_n. If the chosen permutation is written as a product of cycles, then the number of cycles of size i is a random variable $S_i(n)$ for each index $1 \leqslant i \leqslant n$. The distribution of the S_i has been studied since Montmort, who described a game related to the probability that a random permutation has no fixed points; see the references in [3]. The probability that a random injection (defined in Section 2) has no fixed points is derived in Theorem 8 below.

Let $|\pi|$ be the number of cycles of a permutation $\pi \in S_n$. If permutations are chosen proportionally to $\theta^{|\pi|}$, where $\theta > 0$ is a constant, then the distribution of the process of cycle counts is called the Ewens sampling formula [3, 7]. The Ewens sampling formula with $\theta = 1$ is the same as the distribution of cycle counts of random permutations. For the Ewens sampling formula, for each fixed $i \geqslant 1$ the distribution of S_i converges weakly to the Poisson(θ/i) distribution and, moreover, the total variation distance between the process of cycle counts $(S_1, S_2, \ldots, S_b)$ for $b = o(n)$ and the independent Poisson process $(Z_1, Z_2, \ldots, Z_b)$ is $o(1)$, where $Z_i \sim$ Poisson(θ/i) are independent; see [1, 3].

The research in this paper is motivated by the observation that permutations are a particular kind of injection for which the domain is the same as the codomain. The other canonical example of an injection is a maximum matching of a complete bipartite graph, for which the domain and codomain are disjoint. Note that the cycles of a permutation $\pi \in S_n$ correspond to cycles in the directed graph with vertices $\{1, 2, \ldots, n\}$ and directed edges $G_\pi = \{(j, \pi(j)) : 1 \leqslant j \leqslant n\}$. It is shown in Section 2 that the directed graphs corresponding to general injections may be decomposed into components in a way similar to the way directed graphs corresponding to permutations are decomposed into cycles. The components of injections are of two types: paths and cycles. It will be convenient to classify paths themselves into different types. After random injections are defined, limiting properties of the counts of components of the same size and type are derived.

Random injections as defined in this paper do not seem to have been studied before. Random injections have been viewed as a saturated (maximum) matching of a complete bipartite graph in [13]. The components of such an injection are simply vertex disjoint directed edges and elements of the codomain not in the range. The random injections studied here are random digraphs with all of their vertices falling into three predetermined categories: vertices with indegree 0 and outdegree 1; vertices with indegree at most 1 and outdegree 1; and vertices with indegree at most 1 and outdegree 0. Previous research [5] on random digraphs imposing a condition on both indegree and outdegree assumes they both equal a constant common to all vertices.

Components of injections are classified and consequences of the classification are derived in Section 2. We mainly study cycle and path counts. Results on the number of cycle counts of random injections are obtained in Section 3 and path counts are examined in Section 4.

2 The components of injections

Consider an injection f from a finite labelled domain D to a finite labelled codomain R. To each such mapping we associate a directed graph G_f which has vertices $D \cup R$ and directed edges $\{(d, f(d)) : d \in D\} \subseteq D \times R$. If $R = D$, then f is a permutation of D and G_f consists of vertex disjoint directed cycles. On the other hand, if D and R are disjoint, then the injection f is simply a random matching of all elements of D to a subset of R and the graph G_f consists of vertex disjoint directed edges.

We study an interpolation between these two situations. Let n_1, n_2, and n_3 be non-negative integers such that $n_1 + n_2 + n_3 > 0$ and $n_2 \leqslant n_3$. For any mutually disjoint sets A, B and C of sizes $|A| = n_1$, $|B| = n_2$, and $|C| = n_3$, consider the set of injections with domain $A \cup B$ and codomain $A \cup C$. We will call any injection with domain and codomain defined in this way an (n_1, n_2, n_3)-injection. Any injection with finite domain and codomain is an (n_1, n_2, n_3)-injection for some n_1, n_2, n_3. If $n_1 = 0$, then the domain and codomain are disjoint, while if $n_2 = n_3 = 0$, then they are equal. Now that (n_1, n_2, n_3)-injections have been defined, they will usually just be referred to as injections.

The digraph G_f corresponding to an injection f may be decomposed in the following way. A vertex $c \in C$ may not be in the range of f, in which case it may be considered as consisting of an A-path of length 0. We will call such vertices *isolated C vertices*. (Elements of A are called *fixed points* if they map to themselves and may also be thought of as being isolated.) If $c \in C$ is in the range of f, then there are two possibilities; either

c terminates a path of the form $a_1 \to a_2 \to \cdots \to a_\ell \to c$ where $\ell \geqslant 1$ and $a_i \in A$ for all $i \in [1,\ell]$, or c terminates a directed path of the form $b \mapsto a_1 \to a_2 \to \cdots \to a_\ell \to c$ where $\ell \geqslant 0$, $a_i \in A$ for all $i \in [1,\ell]$, and $b \in B$. Let us call a directed path of the first type an *A-path* and a path of the second type an *B-path*. There are always exactly n_2 B-paths and each of them terminates in a unique element of C. Consequently, if $n_2 = n_3$, then there can be no A-paths. Let A_1 denote the set of elements of A which are on A-paths or B-paths. Vertices in the set $A_2 = A \setminus A_1$ must map to elements of $A \cup C$, but they can not map to elements of $A_1 \cup C$ or else they would lie in A_1. Therefore, elements of A_2 map to elements of A_2 and, since f is an injection, the restriction $f|_{A_2}$ is a permutation of A_2. As was noted above, the components of $G_{f|_{A_2}}$ are disjoint directed cycles. We have shown

Lemma 1. *For any injection f, the directed graph G_f is the disjoint union of isolated C vertices, A-paths, B-paths, and cycles.*

We call the isolated C vertices, A-paths, B-paths, and cycles of the lemma the *components* of G_f; the cycles are strongly connected digraph components and the paths are weakly connected digraph components. The *size* of a component is the number of its vertices. The size of any cycle lies between 1 and n_1, the size of any A-path lies between 2 and $n_1 + 1$, and the size of any B-path lies between 2 and $n_1 + 2$.

Lemma 1 is illustrated in Figure 1 in which $A = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$, $B = \{10, 11\}$, $C = \{12, 13, 14, 15, 16\}$ and the injection f is defined by $1 \mapsto 3$, $2 \mapsto 7$, $3 \mapsto 9$, $4 \mapsto 2$, $5 \mapsto 5$, $6 \mapsto 14$, $7 \mapsto 13$, $8 \mapsto 12$, $9 \mapsto 1$, $10 \mapsto 6$, $11 \mapsto 16$. The components of the directed graph G_f are a single isolated C vertex; two cycles whose sizes are 1 and 3; two A-paths whose sizes are 2 and 4; and two B-paths whose sizes are 2 and 3.

Given an injection, let r denote the number of isolated C vertices; let s_i, $1 \leqslant i \leqslant n_1$, denote the number of cycles of size i; let t_j, $2 \leqslant j \leqslant n_1 + 1$, denote the number of A-paths of size j; and let u_k, $2 \leqslant k \leqslant n_1 + 2$, denote the number of B-paths of size j. The total number of vertices is

$$r + \sum_{i=1}^{n_1} i s_i + \sum_{j=2}^{n_1+1} j t_j + \sum_{k=2}^{n_1+2} k u_k = n_1 + n_2 + n_3.$$

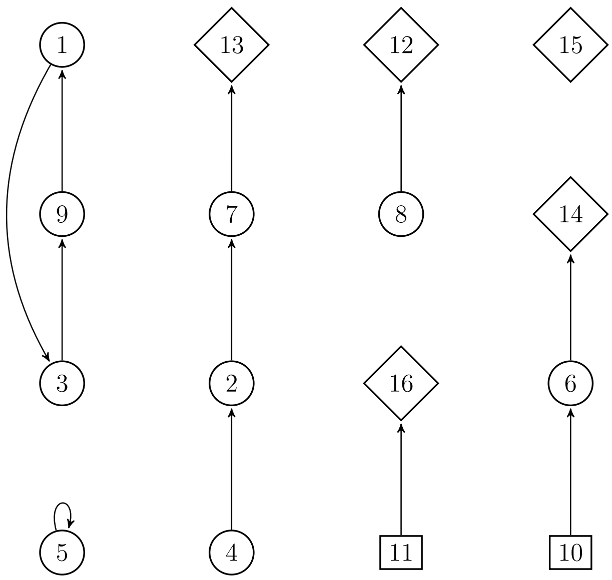
The counts of the A, B and C vertices in the components of an injection must add up to n_1, n_2, and n_3, respectively. Therefore,

$$\sum_{i=1}^{n_1} i s_i + \sum_{j=2}^{n_1+1} (j-1) t_j + \sum_{k=2}^{n_1+2} (k-2) u_k = n_1, \tag{1}$$

$$\sum_{k=2}^{n_1+2} u_k = n_2, \tag{2}$$

and

$$r + \sum_{j=2}^{n_1+1} t_j + \sum_{k=2}^{n_1+2} u_k = n_3. \tag{3}$$



$\bigcirc \in A$, $\Box \in B$, $\Diamond \in C$

Figure 1: The components of a $(9, 2, 5)$-injection

Proposition 2. *Let r be a non-negative integer and let s_i, t_j, and u_k, be sequences of non-negative integers such that (1), (2) and (3) are satisfied. Then the number of injections with exactly r isolated C vertices, s_i cycles of size i, t_j A-paths of size j, and u_k B-paths of size k for $1 \leqslant i \leqslant n_1$, $2 \leqslant j \leqslant n_1 + 1$, $2 \leqslant k \leqslant n_1 + 2$ is*

$$\frac{n_1!n_2!n_3!}{r!\prod_{i=1}^{n_1} s_i!i^{s_i} \prod_{j=2}^{n_1+1} t_j! \prod_{k=2}^{n_1+2} u_k!} \tag{4}$$

Proof. Consider assigning labels to an unlabelled injection with components counts given by r and the s_i, t_j and u_k. There are $n_1!n_2!n_3!$ ways to do this. The injections we obtain in this way all have the desired component structure, but they are not all different. The isolated C vertices can be permuted among each other without changing the resulting injection, as can the vertices forming cycles of the same size, the vertices forming A-paths of the same size, and the vertices forming B-paths of the same size. This accounts for dividing by $r!\prod_{i=1}^{n_1} s_i! \prod_{j=2}^{n_1+1} t_j! \prod_{k=2}^{n_1+2} u_k!$. Finally, each cycle of size i can be obtained in i different ways as any one of its i entries can be in a fixed position of an unlabelled i-cycle, which accounts for dividing by $\prod_{i=1}^{n_1} i^{s_i}$. $\square$

Restricting to $n_2 = n_3 = 0$ gives the following well known formula for the number of permutations with a given cycle structure.

Corollary 3 (Cauchy's Formula). *Let s_i, $i = 1, \ldots, n$ be non-negative integers such that $\sum_{i=1}^{n} is_i = n$. Then the number of permutations on n vertices with exactly s_i cycles of size i, $i = 1, 2, \ldots, n$, is*

$$\frac{n!}{\prod_{i=1}^{n} s_i!i^{s_i}}.$$

A proof of this formula is given on page 92 of [6].

The number of injections is given by $(n_1+n_3)_{n_1+n_2}$, where $(x)_k = x(x-1)\cdots(x-k+1)$ denotes the falling factorial for natural numbers x and k. We will give each injection the uniform measure $1/(n_1+n_3)_{n_1+n_2}$. Our aim is to study the transition from random permutations when $n_2 = n_3 = 0$ to random matchings when $n_1 = 0$.

We define R to be the number of isolated C vertices in a random injection; S_i to be the number of cycles of size i; T_j to be the number of A-paths of size j; and U_k to be the number of B-paths of size k. The joint distributions of these variables can be represented by conditioned independent Poisson variables in the following way. Let $Z^{(1)} \sim \text{Poisson}(1)$, $Z_i^{(2)} \sim \text{Poisson}(1/i)$, $1 \leqslant i \leqslant n_1$, $Z_j^{(3)} \sim \text{Poisson}(1)$, $2 \leqslant j \leqslant n_1+1$, and $Z_k^{(4)} \sim \text{Poisson}(1)$, $2 \leqslant k \leqslant n_1+2$, be mutually independent random variables, and, motivated by (1), (2) and (3), define random variables W_1, W_2 and W_3 by

$$W_1 = \sum_{i=1}^{n_1} iZ_i^{(2)} + \sum_{j=2}^{n_1+1}(j-1)Z_j^{(3)} + \sum_{k=2}^{n_1+2}(k-2)Z_k^{(4)},$$

$$W_2 = \sum_{k=2}^{n_1+2} Z_k^{(4)}$$

and

$$W_3 = Z^{(1)} + \sum_{j=2}^{n_1+1} Z_j^{(3)} + \sum_{k=2}^{n_1+2} Z_k^{(4)}.$$

For any r, s_i, t_j and u_k, let E and E' be the events

$$E = \{R = r\} \cap \bigcap_{i=1}^{n_1}\{S_i = s_i\} \cap \bigcap_{j=2}^{n_1+1}\{T_j = t_j\} \cap \bigcap_{k=2}^{n_1+2}\{U_k = u_k\}$$

and

$$E' = \{Z^{(1)} = r\} \cap \bigcap_{i=1}^{n_1}\{Z_i^{(2)} = s_i\} \cap \bigcap_{j=2}^{n_1+1}\{Z_j^{(3)} = t_j\} \cap \bigcap_{k=2}^{n_1+2}\{Z_k^{(4)} = u_k\}.$$

It can be shown using Proposition 2 that

$$\mathbb{P}(E) = \mathbb{P}(E'|W_1 = n_1, W_2 = n_2, W_3 = n_3). \tag{5}$$

The formula (5) is similar to Theorem 1 of [2] which gives the distribution of random combinatorial structures as independent random variables conditioned on the event that a weighted sum of them equals the size of the object. The formula (5) takes the form

$$\mathbb{P}\left(\bigcap_{i=1}^{n_1}\{S_i = s_i\}\right) = \mathbb{P}\left(\bigcap_{i=1}^{n_1}\{Z_i^{(2)} = s_i\}\Big|\sum_{i=1}^{n_1} iZ_i^{(2)} = n_1\right) \tag{6}$$

when $n_2 = n_3 = 0$. Poisson process approximations of $(S_1, \ldots, S_b)$ for $b = o(n_1)$ were made for random permutations with the assistance of this equality in [1]. Permutations are viewed as a subspecies of the species of endofunctions (also called mappings) in [10]. The component structure of random combinatorial objects such as permutations and mappings are studied in great detail in [3]. For work on the related subject of random set partitions see [4, 8, 14].

Conditional on $|A_2|$, the distribution of the process of cycle counts $(S_1, S_2, \ldots, S_{|A_2|})$ is the distribution of the process of cycle counts of a randomly chosen permutation from $S_{|A_2|}$. Unfortunately, the distribution of $|A_2|$ seems difficult to determine. We will not make use of (5) in this paper. Inclusion-exclusion, the method of moments and the second moment method will be applied instead.

The expected number of isolated C vertices is

$$\mathbb{E}(R) = n_3 \frac{(n_1 + n_3 - 1)_{n_1+n_2}}{(n_1 + n_3)_{n_1+n_2}} = \frac{n_3(n_3 - n_2)}{n_1 + n_3}. \tag{7}$$

If $n_2 = n_3$, then $R = 0$ because all elements of C end a B-path.

In the sequel we consider counts of cycles and paths of different sizes. We let n_2 and n_3 be functions of n_1.

3 The cycles of random injections

The probability mass function of $|A_2|$ will now be derived.

Theorem 4. *If $n_3 \geqslant 1$, then the probability mass function of $|A_2|$ is given by*

$$\mathbb{P}(|A_2| = m) = \frac{n_1!(n_1 + n_3 - m - 1)_{n_3-1}(n_3)_{n_2}}{(n_3 - 1)!(n_1 + n_3)_{n_1+n_2}}, \quad 0 \leqslant m \leqslant n_1. \tag{8}$$

and its cummulative distribution function $|A_2|$ restricted to its support is

$$\mathbb{P}(|A_2| \leqslant m) = 1 - \frac{(n_1)_{m+1}}{(n_1 + n_3)_{m+1}}, \quad 0 \leqslant m \leqslant n_1. \tag{9}$$

Proof. The digraphs of injections with a given $A_2 \subseteq A$, $|A_2| = m$, can be decomposed uniquely into the digraph of a permutation on A_2 together with n_2 B-paths and $n_3 - n_2$ possibly empty A-paths on $A_1 \cup B \cup C$. The number of ways of choosing the paths equals the number of ways of partitioning A_1 into n_3 possibly empty paths, assigning an element of C to each path, and assigning each of n_2 of the paths a single element of B in $(n_3)_{n_2}$ ways. Let D be a set of $n_3 - 1$ labelled elements with D disjoint from $A \cup B \cup C$. Given a permutation of $A_1 \cup D$ written as a sequence and forgetting the labels of D, the elements of D divide A_1, $|A_1| = n_1 - m$, into a sequence of n_3 paths which can be successively assigned elements of C. There are $\frac{(n_1-m+n_3-1)!}{(n_3-1)!}$ ways of doing this. We have shown that there are

$$\binom{n_1}{m} m! \frac{(n_1 + n_3 - m - 1)!}{(n_3 - 1)!} (n_3)_{n_2} = \frac{n_1!(n_1 + n_3 - m - 1)_{n_3-1}(n_3)_{n_2}}{(n_3 - 1)!}$$

injections for which $|A_1| = m$. Dividing by the total number of injections $(n_1+n_3)_{n_1+n_2}$ gives (8). Moreover,

$$\begin{aligned}
\mathbb{P}(|A_2| \leqslant m) &= \sum_{k=0}^{m} \frac{n_1!(n_1+n_3-k-1)_{n_3-1}(n_3)_{n_2}}{(n_3-1)!(n_1+n_3)_{n_1+n_2}} \\
&= \frac{n_1!(n_3)_{n_2}}{(n_3-1)!(n_1+n_3)_{n_1+n_2}} \sum_{j=n_1+n_3-m-1}^{n_1+n_3-1} (j)_{n_3-1} \\
&= \frac{n_1!(n_3)_{n_2}}{n_3!(n_1+n_3)_{n_1+n_2}} \left((n_1+n_3)_{n_3} - (n_1+n_3-m-1)_{n_3}\right) \qquad (10) \\
&= \frac{n_1!}{(n_1+n_3)!} \left((n_1+n_3)_{n_3} - (n_1+n_3-m-1)_{n_3}\right) \\
&= 1 - \frac{(n_1)_{m+1}}{(n_1+n_3)_{m+1}}. \qquad (11)
\end{aligned}$$

where an identity from finite calculus (see [12]) is used at (10) □

Corollary 5.

$$\mathbb{E}|A_2| = \frac{n_1}{n_3+1}. \qquad (12)$$

Therefore, $|A_2| \xrightarrow{P} 0$ *if* $n_1 = o(n_3)$, *where* $\xrightarrow{P}$ *denotes convergence in probability as* $n_1 \to \infty$.

Proof.

$$\begin{aligned}
\mathbb{E}|A_2| &= \sum_{m=1}^{n_1} \mathbb{P}(|A_2| \geqslant m) \\
&= \sum_{m=1}^{n_1} \frac{(n_1)_m}{(n_1+n_3)_m} \\
&= \frac{n_1!}{(n_1+n_3)!} \sum_{m=1}^{n_1} (n_1+n_3-m)_{n_3} \\
&= \frac{n_1!}{(n_1+n_3)!} \frac{(n_1+n_3)_{n_3+1}}{n_3+1} \\
&= \frac{n_1}{n_3+1}.
\end{aligned}$$

□

Assuming $n_3 = o(n_1)$, the next lemma shows that $|A_2| \to \infty$ as $n_1 \to \infty$ with high probability.

Lemma 6. *If* $n_3 = o(n_1)$, *then there is an integer valued function* $\omega(n_1)$ *for which* $\omega(n_1) \to \infty$ *as* $n_1 \to \infty$ *and such that*

$$\mathbb{P}(|A_2| \geqslant \omega(n_1)) = 1 - o(1), \quad n_1 \to \infty.$$

Proof. For ω growing slowly enough we have

$$\begin{aligned}
\mathbb{P}(|A_2| \geqslant \omega) &= \frac{(n_1)_\omega}{(n_1+n_3)_\omega} \\
&\geqslant \left(\frac{n_1-\omega}{n_1+n_3-\omega}\right)^\omega \\
&= \exp\left(\omega \log\left(1-\frac{\omega}{n_1}\right) - \omega\log\left(1+\frac{n_3-\omega}{n_1}\right)\right) \\
&= \exp\left(O\left(\max(\omega^2/n_1, \omega n_3/n_1)\right)\right).
\end{aligned}$$

Define

$$\omega(n_1) = \left\lfloor \min\left(n_1^{1/4}, (n_1/n_3)^{1/2}\right) \right\rfloor,$$

where $\lfloor x \rfloor$ is the integer part of x. □

Let $C_i(n)$ denote the number of cycles of length i in a random permutation on n letters and let $S_i(n_1)$ denote the number of cyces of size i in a random injection. We can use Lemma 6 to transfer results on the process $(C_1(n), C_2(n), \ldots)$ to $(S_1(n_1), S_2(n_1), \ldots)$. The total variation distance between the laws $\mathcal{L}(X)$ and $\mathcal{L}(Y)$ of random elements taking values in a discrete space $\mathcal{S}$ is defined to be

$$d_{\mathrm{TV}}(\mathcal{L}(X), \mathcal{L}(Y)) = \sup_{A\subseteq\mathcal{S}}(\mathbb{P}(X\in A) - \mathbb{P}(Y \in \mathcal{A})).$$

It is shown in [1] that for $1 \leqslant b \leqslant n$,

$$d_{\mathrm{TV}}(\mathcal{L}(C_1(n), C_2(n), \ldots, C_b(n)), \mathcal{L}(Z_1, Z_2, \ldots, Z_b)) \leqslant F(n/b),$$

where the Z_i are independent Poisson distributed random variables with parameters $\lambda_i = 1/i$, for an explicit function satisfying $\log F(x) \sim -x\log x$ as $x \to \infty$.

Theorem 7. *If $b(n_1) = o(\omega(n_1))$, then*

$$d_{\mathrm{TV}}(\mathcal{L}(S_1(n_1), S_2(n_1), \ldots, S_b(n_1)), \mathcal{L}(Z_1, Z_2, \ldots, Z_b)) = o(1),$$

Proof. For n_1 large enough so that $b(n_1) \leqslant \omega(n_1)$, We have

$$\begin{aligned}
& d_{\mathrm{TV}}(\mathcal{L}(S_1(n_1), S_2(n_1), \ldots, S_b(n_1)), \mathcal{L}(Z_1, Z_2, \ldots, Z_b)) \\
\leqslant\ & \sum_{n=0}^{n_1} \mathbb{P}(|A_2| = n)\, d_{\mathrm{TV}}(\mathcal{L}(C_1(n), C_2(n), \ldots, C_b(n)), \mathcal{L}(Z_1, Z_2, \ldots, Z_b)) \\
\leqslant\ & \mathbb{P}(|A_2| \leqslant \omega(n_1)) + \sum_{n=\omega(n_1)+1}^{n_1} \mathbb{P}(|A_2| = n) F(n/b) \\
\leqslant\ & \mathbb{P}(|A_2| \leqslant \omega(n_1) + \max_{\omega(n_1)+1\leqslant n\leqslant n_1} F(n/b) \\
=\ & o(1).
\end{aligned}$$

□

Let $L_i(n_1)$ be the ith largest cycle in a random injection. Lemma 6 and a limit result on the large cycles of a random permutation from [11, 15] can be used to show

$$|A_2|^{-1}(L_1(n_1), L_2(n_1), \ldots) \xrightarrow{d} (L_1, L_2, \ldots) \quad \text{in } \mathbb{R}^\infty$$

where the left hand side is taken arbitrarily to be $(1, 0, 0, \ldots)$ if $A_2 = \emptyset$, the distribution of the random vector $(L_1, L_2, \ldots)$ is the Poisson-Dirichlet distribution with parameter $\theta = 1$, and the convergence is in distribution. For various representations of the Poisson-Dirichlet distribution see [3].

Inclusion-exclusion can be used to find the exact probability that a random injection has no fixed points

Theorem 8. *The probability that a random injection has no fixed points is*

$$\mathbb{P}(S_1 = 0) = \sum_{l=0}^{n_1} \frac{(-1)^l}{l!} \frac{(n_1)_l}{(n_1 + n_3)_l}. \tag{13}$$

Proof. Let $F \subseteq A$. The injections f for which F are fixed points are those such that $f|_F$ is the identity map and $f|_{(A \setminus F) \cup B}$ is an injection with codomain $(A \setminus F) \cup C$. Therefore,

$$\mathbb{P}(E_F) = \frac{(n_1 - |F| + n_3)_{n_1 - |F| + n_2}}{(n_1 + n_3)_{n_1 + n_2}} = \frac{1}{(n_1 + n_3)_{|F|}}.$$

It follows from inclusion-exclusion that

$$\mathbb{P}(S_1 = 0) = \sum_{l=0}^{n_1} (-1)^l \binom{n_1}{l} \frac{1}{(n_1 + n_3)_l} = \sum_{l=0}^{n_1} \frac{(-1)^l}{l!} \frac{(n_1)_l}{(n_1 + n_3)_l}.$$

□

Formula (13) shows that the probability of having no fixed points does not depend at all on n_2.

The next theorem shows that the counts of cycles of different sizes exhibit a phase transition when the size of n_3 is of the same order as n_1. Given sequences x_{n_1}, y_{n_1}, we write $x_{n_1} \sim y_{n_1}$ to mean $\lim_{n_1 \to \infty} x_{n_1}/y_{n_1} = 1$.

Theorem 9. *The expected number of cycles of size* $1 \leqslant i \leqslant n_1$ *is*

$$\mathbb{E}(S_i) = \frac{1}{i} \frac{(n_1)_i}{(n_1 + n_3)_i}. \tag{14}$$

Suppose that $n_3 \sim \gamma n_1$ *for a constant* $\gamma > 0$. *For each fixed* $d \geqslant 1$, *the process of small cycles satisfies*

$$(S_1, S_2, \ldots, S_d) \xrightarrow{D} (Z_1, Z_2, \ldots, Z_d) \tag{15}$$

Z_i *are mutually independent* Poisson(λ_i) *distributed random variables with parameters*

$$\lambda_i = \frac{1}{i}(1 + \gamma)^{-i} \tag{16}$$

If $\omega(n_1)$ is any integer valued function growing to infinity arbitrarily slowly, the number of vertices on cycles of size at least $\omega(n_1)$ satisfies

$$\sum_{i=\omega(n_1)}^{n_1} iS_i \xrightarrow{P} 0. \tag{17}$$

Proof. We will calculate the joint falling moments of the S_i. For any $\mu_i \geqslant 0$, $i = 1, \dots, d$, let $\Gamma(\mu_1, \dots, \mu_d)$ denote the set of sequences of $\sum_{i=1}^d \mu_i$ vertex disjoint cycles such that the first μ_1 of them have size 1, the next μ_2 of them have size 2, and so on until the last μ_d of them have size d. Define $\tau = \sum_{i=1}^d i\mu_i$. For $\tau \leqslant n_1$, the size of Γ is

$$\begin{aligned} |\Gamma(\mu_1, \dots, \mu_d)| &= \frac{n_1!}{1!^{\mu_1} 2!^{\mu_2} \cdots d!^{\mu_d} (n_1 - \tau)!} \prod_{i=1}^{d} (i-1)!^{\mu_i} \\ &= \frac{n_1!}{1^{\mu_1} 2^{\mu_2} \cdots d^{\mu_d} (n_1 - \tau)!}. \end{aligned}$$

Given $(\alpha_1, \dots, \alpha_{\sum_{i=1}^d \mu_i}) \in \Gamma(\mu_1, \dots, \mu_d)$, let $A_l \subseteq A$ be the set of vertices in cycle α_l, $l = 1, \dots, \sum_{i=1}^d \mu_i$, and define

$$F = \bigcup_{l=1}^{\sum_{i=1}^d \mu_i} A_l.$$

The injections for which all α_l are components are precisely those for which $f|_F$ is determined by the α_l and $f|_{(A \setminus F) \cup B}$ is an injection with codomain $(A \setminus F) \cup C$. Note that $|F| = \tau$. The joint falling moment of the S_i corresponding to the μ_i is

$$\begin{aligned} \mathbb{E}((S_1)_{\mu_1} \cdots (S_d)_{\mu_d}) &= \frac{n_1!}{1^{\mu_1} 2^{\mu_2} \cdots d^{\mu_d} (n_1 - \tau)!} \frac{(n_1 - \tau + n_3)_{n_1 - \tau + n_2}}{(n_1 + n_3)_{n_1 + n_2}} \\ &= \frac{1}{1^{\mu_1} 2^{\mu_2} \cdots d^{\mu_d}} \frac{(n_1)_\tau}{(n_1 + n_3)_\tau}. \end{aligned}$$

In particular, the formula (14) follows by taking $\mu_i = 1$ and $\mu_m = 0$ for $m \neq i$ for each $1 \leqslant i \leqslant n_1$.

As $n_1 \to \infty$,

$$\mathbb{E}((S_1)_{\mu_1} \cdots (S_d)_{\mu_d}) \sim \prod_{i=1}^{d} \lambda_i^{\mu_i}.$$

The conclusion of weak convergence (15) follows from the method of moments (Theorem 6.2 of [9]) with λ_i given by (16) if $n_3 \sim \gamma n_1$ and $\lambda_i = 1/i$ if $n_3 = o(n_1)$ as in Theorem 7.

By (14), the expected number of vertices on cycles of size at least $\omega(n_1)$ is

$$\begin{aligned}\mathbb{E}\left(\sum_{i=\omega(n_1)}^{n_1} iS_i\right) &= \frac{n_1!}{(n_1+n_3)!}\sum_{i=\omega(n_1)}^{n_1}(n_1+n_3-i)_{n_3} \\ &= \frac{n_1!}{(n_1+n_3)!}\sum_{j=1}^{n_1+n_3-\omega(n_1)}(j)_{n_3} \\ &= \frac{n_1!}{(n_1+n_3)!}\frac{(n_1+n_3-\omega(n_1)+1)_{n_3+1}}{n_3+1} \\ &= \frac{(n_1)_{\omega(n)}}{(n_1+n_3)_{\omega(n)-1}}\frac{1}{n_3+1},\end{aligned}$$

which converges to 0 as $n_1 \to \infty$ when $n_3 \sim \gamma n_1$, implying (17). $\square$

The limiting distribution in (15) is the same as for the Ewens sampling formula with $\theta = (1+\gamma)^{-1}$. However, unlike the Ewens sampling formula, asymptotically a random injection does not have any cycles of size growing to infinity as $n_1 \to \infty$.

4 Counts of paths of different sizes

In this section we estimate the number of A-paths and B-paths of different sizes and get almost sure asymptotics for the maximum size of the two kinds of paths under certain conditions.

Theorem 10. *Suppose $n_3 \geqslant 1$. The expectation of the number of A-paths of size $2 \leqslant i \leqslant n_1+1$ is*

$$\mathbb{E}(T_i) = \frac{(n_1)_{i-1}n_3(n_3-n_2)}{(n_1+n_3)_i} \tag{18}$$

and the expectation of the number of B-paths of size $2 \leqslant i \leqslant n_1+2$ is

$$\mathbb{E}(U_i) = \frac{(n_1)_{i-2}n_2n_3}{(n_1+n_3)_{i-1}}.$$

Let the sequence of indices $i = i(n_1)$ be such that $i = o(\sqrt{n_1})$. If $\lim_{n_1\to\infty}\min(n_2, n_3-n_2) \to \infty$ and $\lim_{n_1\to\infty}\mathbb{E}(T_i) = \infty$, then

$$\frac{T_i}{\mathbb{E}(T_i)} \xrightarrow{P} 1; \tag{19}$$

while if $\lim_{n_1\to\infty} n_2 \to \infty$ and $\lim_{n_1\to\infty}\mathbb{E}(U_i) = \infty$, then

$$\frac{U_i}{\mathbb{E}(U_i)} \xrightarrow{P} 1.$$

Proof. For each fixed $1 \leqslant \mu \leqslant \min(n_1/(i-1), n_3)$, the number of ways of choosing μ ordered vertex disjoint A-paths of size i is $(n_1)_{\mu(i-1)}(n_3)_\mu$. Let α denote a particular set of μ ordered vertex disjoint A-paths of size i. Let F be the subset of A which are vertices of A-paths in α and let G be the subset of C which are vertices of A-paths in α, where $|F| = \mu(i-1)$ and $|G| = \mu$. The injections f for which the chosen A-paths are components are those for which $f|_F$ is determined by α and $f|_{(A\setminus F)\cup B}$ is an injection with codomain $(A \setminus F) \cup (C \setminus G)$. Therefore, the μth falling factorial moment of T_i is

$$\begin{aligned}\mathbb{E}(T_i)_\mu &= (n_1)_{\mu(i-1)}(n_3)_\mu \frac{(n_1 - \mu(i-1) + n_3 - \mu)_{n_1-\mu(i-1)+n_2}}{(n_1+n_3)_{n_1+n_2}} \\ &= \frac{(n_1)_{\mu(i-1)}(n_3)_\mu(n_3-n_2)_\mu}{(n_1+n_3)_{\mu i}},\end{aligned}$$

Similarly, for each $\mu > 0$ such that $\mu(i-2) \leqslant n_1$ and $\mu \leqslant n_2$, the μth falling factorial moment of the number of B-paths of size i is

$$\begin{aligned}\mathbb{E}(U_i)_\mu &= (n_1)_{\mu(i-2)}(n_2)_\mu(n_3)_\mu \frac{(n_1 - \mu(i-2) + n_3 - \mu)_{n_1-\mu(i-2)+n_2-\mu}}{(n_1+n_3)_{n_1+n_2}} \\ &= \frac{(n_1)_{\mu(i-2)}(n_2)_\mu(n_3)_\mu}{(n_1+n_3)_{\mu(i-1)}}.\end{aligned}$$

Letting $\mu = 1$ in the formulae for the falling moments gives the formulae for $\mathbb{E}(T_i)$ and $\mathbb{E}(U_i)$.

By the assumptions $\lim_{n_1\to\infty} \min(n_2, n_3 - n_2) \to \infty$ and $i = o(\sqrt{n_1})$, letting $\mu = 2$ in the formulae for the falling moments results in

$$\mathbb{E}(T_i)_2 \sim \left(\frac{n_1^{i-1} n_3(n_3-n_2)}{(n_1+n_3)^i}\right)^2 \sim (\mathbb{E}(T_i))^2,$$

while the assumptions $\lim_{n_1\to\infty} n_2 \to \infty$, $n_2 \leqslant n_3$, and $i_1 = o(\sqrt{n_1})$ produce

$$\mathbb{E}(U_i)_2 \sim \left(\frac{n_1^{i-2} n_2 n_3}{(n_1+n_3)^{i-1}}\right)^2 \sim (\mathbb{E}(U_i))^2.$$

The identity $\mathrm{Var}(X) = \mathbb{E}(X)_2 + \mathbb{E}(X) - (\mathbb{E}(X))^2$, true for any random variable X, and $\lim_{n_1\to\infty} \mathbb{E}(T_i) = \infty$ and $\lim_{n_1\to\infty} \mathbb{E}(U_i) = \infty$ now gives us the estimates $\mathrm{Var}(T_i) = o((\mathbb{E}(T_i)^2)$ and $\mathrm{Var}(U_i) = o((\mathbb{E}(U_i)^2)$ The conclusions about convergence in probability result from the second moment method; see [9], for example. □

Define

$$Y_A = \max\{i \geqslant 2 : T_i > 0\}$$

and

$$Y_B = \max\{i \geqslant 2 : U_i > 0\}$$

to be the maximum sizes of A-paths and B-paths, respectively. Under suitable conditions the next theorem provides asymptotics for Y_A and Y_B.

Theorem 11. *Given a function $\omega(n_1)$ which converges to infinity arbitrarily slowly, define*

$$i_A = 1 + \left\lfloor \frac{\log\left(\frac{n_3(n_3-n_2)}{n_1+n_3}\right) - \omega(n_1)}{\log\left(\frac{n_1+n_3}{n_1}\right)} \right\rfloor, \quad j_A = 1 + \left\lceil \frac{\log\left(\frac{n_3(n_3-n_2)(n_1+n_3-1)}{(n_1+n_3)(n_3-1)}\right) + \omega(n_1)}{\log\left(\frac{n_1+n_3-1}{n_1}\right)} \right\rceil,$$

where for any real number x, $\lfloor x \rfloor$ and $\lceil x \rceil$ are the usual floor and ceiling functions. If $\lim_{n_1\to\infty} \min(n_2, n_3 - n_2) \to \infty$, $\liminf_{n_1\to\infty} i_A \geqslant 2$, and $i_A = o(\sqrt{n_1})$, then

$$\lim_{n_1\to\infty} \mathbb{P}\left(i_A \leqslant Y_A < j_A\right) = 1. \tag{20}$$

Define

$$i_B = 2 + \left\lfloor \frac{\log\left(\frac{n_2 n_3}{n_1+n_3}\right) - \omega(n_1)}{\log\left(\frac{n_1+n_3}{n_1}\right)} \right\rfloor, \quad j_B = 2 + \left\lceil \frac{\log\left(\frac{n_2 n_3(n_1+n_3-1)}{(n_1+n_3)(n_3-1)}\right) + \omega(n_1)}{\log\left(\frac{n_1+n_3-1}{n_1}\right)} \right\rceil$$

If $\lim_{n_1\to\infty} n_2 \to \infty$, $\liminf_{n_1\to\infty} i_B \geqslant 2$, and $i_B = o(\sqrt{n_1})$, then

$$\lim_{n_1\to\infty} \mathbb{P}\left(i_B \leqslant Y_B < j_B\right) = 1. \tag{21}$$

Proof. We prove the theorem first for Y_A. The lower bound on i_A and (18) imply that by (18), for any $i_A \leqslant i \leqslant n_1 + 2$,

$$\mathbb{E}(T_i) = \frac{(n_1)_{i-1} n_3(n_3-n_2)}{(n_1+n_3-1)_{i-1}(n_1+n_3)} \leqslant \left(\frac{n_1}{n_1+n_3-1}\right)^{i-1} \frac{n_3(n_3-n_2)}{n_1+n_3}$$

and so by the definition of j_A,

$$\begin{aligned} \mathbb{P}\left(\bigcup_{i\geqslant j_A} \{T_i > 0\}\right) &\leqslant \sum_{i\geqslant j_A} \mathbb{P}(T_i > 0) \leqslant \sum_{i\geqslant j_A} \mathbb{E}(T_i) \\ &= \left(\frac{n_1}{n_1+n_3-1}\right)^{j_A-1} \frac{n_3(n_3-n_2)(n_1+n_3-1)}{(n_1+n_3)(n_3-1)} = o(1). \end{aligned} \tag{22}$$

By (18) and $i_A = o(\sqrt{n_1})$,

$$\mathbb{E}(T_{i_A}) \sim \frac{n_1^{i_A-1} n_3(n_3-n_2)}{(n_1+n_3)^{i_A}} = \left(\frac{n_1}{n_1+n_3}\right)^{i_A-1} \frac{n_3(n_3-n_2)}{n_1+n_3}$$

and so by the definition of i_A, $\lim_{n_1\to\infty} \mathbb{E}(T_{i_A}) = \infty$. This fact along with the hypotheses of the theorem imply that Theorem 18 can be applied with $i = i_A$. Therefore, (19) results in $\lim_{n_1\to\infty} \mathbb{P}(T_{i_A} > 0) = 1$, hence $\lim_{n_1\to\infty} \mathbb{P}(Y_A \geqslant i_A) = 1$. Together with (22), this proves (20).

The proof of (21) is similar and omitted. □

Applications of Theorem 11 demonstrate how Y_A and Y_B depend on n_1, n_2, and n_3: as n_2 and n_3 are made larger in tandem relative to n_1 the maximal sizes of A-paths and B-paths become smaller. Let γ_2, γ_3 be constants such that $\gamma_3 > \gamma_2 > 0$ for Y_A and $\gamma_3 \geqslant \gamma_2 > 0$ for Y_B. If $n_2 \sim \gamma_2 n_1^\kappa$, $n_3 \sim \gamma_3 n_1^\kappa$, $\kappa \in (1/2, 1)$, then $i_A \sim i_B \sim \frac{2\kappa-1}{\gamma_3} {n_1}^{1-\kappa} \log n_1$ and $j_A \sim j_B \sim \frac{\kappa}{\gamma_3} {n_1}^{1-\kappa} \log n_1$. If $n_2 \sim \gamma_2 n_1$ and $n_3 \sim \gamma_3 n_1$ then $i_A \sim i_B \sim j_A \sim j_B \sim \log n_1 / \log(1+\gamma_3)$. If $n_2 \sim \gamma_2 n^\kappa$, $n_3 \sim \gamma_3 n^\kappa$, $\kappa > 1$, $i_A = 1 + \lfloor \frac{\kappa}{\kappa-1} + o(1) \rfloor$, $j_A = 1 + \lceil \frac{\kappa}{\kappa-1} + o(1) \rceil$, $i_B = 2 + \lfloor \frac{\kappa}{\kappa-1} + o(1) \rfloor$, and $j_B = 2 + \lceil \frac{\kappa}{\kappa-1} + o(1) \rceil$. It follows that $j_A - i_A \leqslant 2$ and $j_B - i_B \leqslant 2$ for n_1 large enough and so Y_A and Y_B have asymptotic two-point concentrations with high probability. It follows that $j_A - i_A \leqslant 2$ and $j_B - i_B \leqslant 2$ for n_1 large enough and so Y_A and Y_B have asymptotic two-point concentrations with high probability. They have asymptotic one-point concentrations if $\frac{\kappa}{\kappa-1}$ is not an integer. For $\kappa > 2$, $Y_A \xrightarrow{P} 2$ and $Y_B \xrightarrow{P} 3$. We have

$$\sum_{i \geqslant 3} \mathbb{E}(U_i) \leqslant \sum_{i \geqslant 3} \left(\frac{n_1}{n_1 + n_3 - 1} \right)^{i-2} \frac{n_2 n_3}{n_1 + n_3} = \frac{n_1 n_2 n_3}{(n_1 + n_3)(n_3 - 1)}$$

and therefore when $n_1 n_2 = o(n_3)$ and $n_2 > 0$ we have $Y_B \xrightarrow{P} 2$.

Acknowledgements

Thanks to the referee for supplying some references.

References

[1] R. Aratia and S. Tavaré, The cycle structure of random permutations. *Ann. Probab.*, 20(3):1567–1591, 1992.

[2] R. Arratia and S. Tavaré, Independent process approximations for random combinatorial structures. *Adv. Math.* 104(1):90-154, 1994.

[3] R. Arratia, A. D. Barbour, and S. Tavaré, Logarithmic combinatorial structures: a probabilistic approach. EMS Monographs in Mathematics. European Mathematical Society, Zürich, 2003.

[4] B. Chern, P. Diaconis, D. M. Kane, and R. C. Rhoades, Central limit theorems for some set partition statistics. *Adv. in Appl. Math.* 70:92-105, 2015.

[5] C. Cooper, A. Frieze, M. Molloy, Hamilton cycles in random regular digraphs. *Combin., Probab. Comput.* 3(1):39-49, 1994.

[6] M. Bóna, Combinatorics of Permutations. Second Edition. Discrete Mathematics and its applications. CRC Press, Boca Raton, 2012.

[7] W. J. Ewens, The sampling theory of selectively neutral alleles. *Theoret. Population Biol.* 3:87–112, 1972.

[8] B. Fristedt, The structure of partitions of large sets. Technical Report 86-154, Mathematics Dept., University of Minnesota, Minneapolis, Minnesota, 1987.

[9] S. Janson, T. Łuczak, and A. Ruciński, Random Graphs. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000.

[10] A. Joyal, Une théorie combinatoire des séries formelles [A combinatorial theory of formal series]. *Adv. in Math.* 42(1):1–82, 1981.

[11] J. F. C. Kingman, The population structure associated with the Ewens sampling formula. *Theoret. Population Biol.* 11(2):274 – 283, 1977.

[12] R. L. Graham, D. E. Knuth, O. Patashnik, Concrete mathematics: A foundation for computer science. Second edition. Addison-Wesley Publishing Company, 1994.

[13] L. Lu and L. Székely, Using Lovász local lemma in the space of random injections. *Electron. J. Combin.* 14(1):13 pp, 2007.

[14] J. Pitman, Some probabilistic aspects of set partitions. *Amer. Math. Monthly* 104(3):201-209, 1997.

[15] A. M. Vershik and A. A. Schmidt, Limit measures arising in the asymptotic theory of symmetric groups. I., *Theory Probab. Appl.* 22(1):70–85, 1977.